

Article

FRIMFL: A Fair and Reliable Incentive Mechanism in Federated Learning

Abrar Ahmed  and Bong Jun Choi * 

School of Computer Science and Engineering, Soongsil University, Seoul 06978, Republic of Korea; abrar@soongsil.ac.kr

* Correspondence: davidchoi@soongsil.ac.kr

Abstract: Federated learning (FL) enables data owners to collaboratively train a machine learning model without revealing their private data and sharing the global models. Reliable and continuous client participation is essential in FL for building a high-quality global model via the aggregation of local updates from clients over many rounds. Incentive mechanisms are needed to encourage client participation, but malicious clients might provide ineffectual updates to receive rewards. Therefore, a fair and reliable incentive mechanism is needed in FL to promote the continuous participation of clients while selecting clients with high-quality data that will benefit the whole system. In this paper, we propose an FL incentive scheme based on the reverse auction and trust reputation to select reliable clients and fairly reward clients that have a limited budget. Reverse auctions provide candidate clients to bid for the task while reputations reflect their trustworthiness and reliability. Our simulation results show that the proposed scheme can accurately select users with positive contributions to the system based on reputation and data quality. Therefore, compared to the existing schemes, the proposed scheme achieves higher economic benefit encouraging higher participation, satisfies reward fairness and accuracy to promote stable FL development.

Keywords: federated learning; client selection; reverse auction; incentive mechanism



Citation: Ahmed, A.; Choi, B.J. FRIMFL: A Fair and Reliable Incentive Mechanism in Federated Learning. *Electronics* **2023**, *12*, 3259. <https://doi.org/10.3390/electronics12153259>

Academic Editors: Weiting Zhang, Chuan Zhang and Tong Wu

Received: 30 May 2023
Revised: 20 July 2023
Accepted: 25 July 2023
Published: 28 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Federated learning is a promising machine learning technique that aims to preserve the privacy of data owners as they collaboratively train a global model without exposing their raw data [1]. Its potential to assist artificial intelligence in achieving unprecedented success is phenomenal in solving user privacy problems and data islands. As shown in Figure 1, each data owner downloads a global model from a server and trains it locally using its private data samples and computational resources. After local data training, the parameters of locally trained models are aggregated with the server's global model [2] using the model averaging (FedAvg) algorithm. This process continues for multiple training rounds until the desired accuracy or objective function is achieved. Since there is no upload of raw data, FL drastically improves data privacy and security issues. FL's salient privacy-preserving feature and numerous high system characteristics have gained attention and research in academia and industry [3], enabling its broad applicability in various fields such as finance, IoT, health care, and telecommunication. For instance, Google adopted FL in its keyboard application GBoard to improve the performance [4].

Despite its numerous advantages, federated learning still faces significant challenges. Most existing studies assume that all data owners participate unconditionally [5] and honestly contribute data [6]. But this assumption is impractical due to communication and computational resource costs incurred by inevitable training [7]. Simultaneously, clients might face additional barriers of privacy threats and information leakages. Therefore, self-interested individuals will hesitate to participate in model training [8] without any incentives and not serve for free [9]. Hence, without any rewards, individuals will not

remain in the FL system, leading to untenable performance. Meanwhile, some data owners might exhibit undesirable behaviors by deceptively training results via submitting poor models intentionally or unintentionally to improve their utilities; the server may need to be fully aware of their computational resources, amount, and data quality. Participants may launch poisoning attacks [10] or malicious updates affecting the global model learning performance. They might also launch free-riding attacks to trick the system and obtain the rewards [11]. Fairness in an emerging trustworthy FL pillar demands a multidisciplinary approach [12,13]. Unfairness exists during client selection [14], model optimization [15], contribution evaluation stages, and incentive distribution [3] that can adversely impact the FL server and clients if not appropriately addressed. Current research in incentive mechanisms lacks fairness in the aggregation phase to weigh individual updates according to their performances [16] and the reward phase in distributing models/payoffs according to their contribution [17]. Clients might be discouraged from enrolling in FL training by unfair treatment. However, equal treatment for all clients without considering their potential contributions can reduce server capability to attract high-quality users, leading to poorly generalized FL models [18]. Moreover, the studies are ineffective in coupling reliable client selection with fair incentive mechanisms, as evaluating client behaviors is inefficient. The clients are inconsistent and have no motivation to participate. Hence, designing an efficient and reliable incentive mechanism in FL is crucial to stimulate clients to contribute quality work and promote stable FL development.

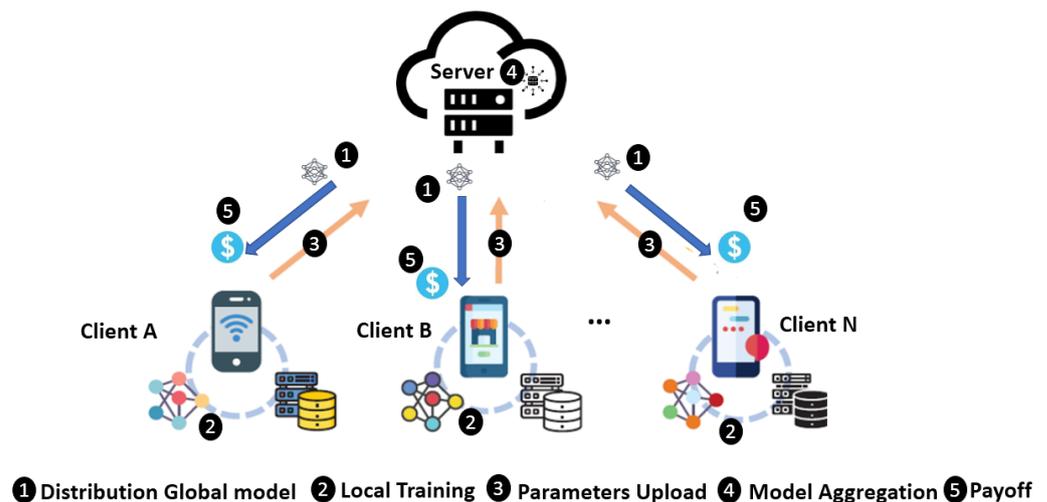


Figure 1. Workflow of incentive distribution in FL.

In this study, we propose an incentive mechanism based on reverse auctions for pricing and reputation trust to encourage clients towards active participation and permit model owners to receive high-quality results. Reverse auctions effectively model training costs and encourage users to participate truthfully. Reputation measures useful updates through trust metrics to reflect an individual's reliability. Client behaviors to manipulate the training process are simulated with at least 60% of malicious poisoning data that lead to slow model convergence and performance in the test environment. To evaluate reputation, we propose a quality-aware selection to examine and differentiate high-quality workers. The server stores and updates candidates' information in the incentive module.

To summarize, our contributions are as follows:

- We proposed a reliable FL incentive scheme (FRIMFL) that combines reverse auction and reputation to incentivize clients.
- We constructed a weighted trust assessment method to reflect clients' reliability considering the quality of model updates.

- We introduced Shapley method to derive the per-round marginal contributions of participants. FRIMFL incorporates reputation (computed from trust and contribution measures) in fair reward allocation to participants.
- The simulation analysis regarding social welfare, contribution fairness, and accuracy shows that our proposed mechanism is compatible, individually rational, and budget feasible.

The remainder of this paper is organized as follows. Section 2 presents related works. Section 3 elaborates on the taxonomy of incentive schemes. Section 4 provides materials and methods consisting of details of the proposed FRIMFL mechanism that distributes incentives using quality, trust, and contribution assessments. Section 5 presents a theoretical analysis to show the economic properties of the proposed mechanism. Section 6 includes the performance evaluation in terms of economic benefits, client selection performance, model accuracy, and fairness. Finally, Section 7 provides discussions, and Section 8 concludes the paper.

2. Related Work

McMahan et al. [1] proposed a FedAvg algorithm that calculates the local model weights accordingly to the client data amount. However, clients may act as free riders or perform malicious attacks like data poisoning attacks. Incentive schemes were proposed to remedy these problems. Client selection, contribution evaluation, and payment allocation are the fundamental building components for incentive mechanisms [19], where optimal reward strategies for both FL servers and clients are computed through objective optimization functions. Kang et al. [20] presented an FL incentive scheme based on a combination of contract theory and reputation using a subjective logic model (SLM). Cong et al. [21] proposed an incentive design in mobile networks using contract theory to tackle the information asymmetry issue for attracting high-quality workers. Similarly, Ye et al. [22] introduced a 2D contract based on client quality and computational capabilities to determine the rewards in a monopolist system. However, the workers' pricing cost for training remains an open question before incentivizing.

Fairness in FL is examined as resource distribution and model optimization using personalization techniques [15]. Ezzeldin et al. [23] proposed a group fairness strategy to mitigate bias in FL models. In [24], a federated dropout pruning approach was presented to customize client models. However, these works only focus on selection and group fairness notions [18]. For context, to deal with fairness from the contribution notion, Zhang et al. [25] measured client contributions based on self-reported information in hierarchical FL. The information includes quality, quantity, communication capabilities, and data collection costs for joining the system. In [26], the authors introduced a mutual evaluation approach where each client receives points based on local credibility protected by differential privacy and blockchain. Michieli and Ozay [16] proposed FairFL for the uniform treatment of users based on their contributions via fair aggregation of model weights. Le et al. [27] adopted auctions using self-reporting where the server evaluates bid information for contribution and winner determination. Cong et al. [28] proposed a Vickrey–Clarke–Grove (VCG) scheme to incentivize participants to report costs and quality truthfully. However, self-reported information does not guarantee truthfulness in practice. Zheng et al. [29] introduced a multi-dimensional auction scheme to motivate high-quality workers to participate. Still, there is no assurance that selected individuals would work according to the bidding agreement.

Nishio et al. [14] proposed an FedCS protocol for high-quality client selection with restricted resources. They overshadowed the reliability and data accuracy aspects by considering only computational and communication costs. Zhang et al. [17] proposed a scheme based on auction and reputation to incentivize clients instead of self-reporting. The reputation depends on the contribution evaluation via cosine similarity between the client and global models. However, this might not be reliable for clients with highly imbalanced distribution. Moreover, the formulation of trust through Gompertz's function as a static non-linear mathematical model does not take into account subjective beliefs. It models the

trust in a smooth sigmoidal curve, which might not always hold in real-world scenarios. Reputation is an essential metric in FL incentive schemes. Rehman et al. [30] presented a reputation-based blockchain system for client selection. However, the proposed scheme lacks a comprehensive consideration of client contributions to model updates.

The above-mentioned notable works drive client selection and reward systems under threshold-based approaches, thus resulting in the disinterest of FL participants. It is crucial to strike a balance between the interests of all entities in the FL system. Therefore, to address the fairness and reliable evaluation of client performance, our scheme (FRIMFL), based on a reputation mechanism with subjective trust quality, federated Shapley contribution, and reverse auctions, promotes quality performance and ensures contribution fairness during both the client selection and contribution evaluation stages.

3. Taxonomy of Incentive Mechanisms

Incentives motivate data owners to continue participating in the federated learning environment. They can be classified as positive and negative incentives. Positive incentives quest to motivate participants by recognizing the positive impact of contributions and encouraging rewards. Negative incentives intend to avoid malicious attackers or negative participation through reputation, penalty loss, etc. Current studies to incentivize workers can be classified into settings, stages, methods, and research challenges, as described in the following subsections. Figure 2 illustrates the taxonomy of incentive mechanisms.

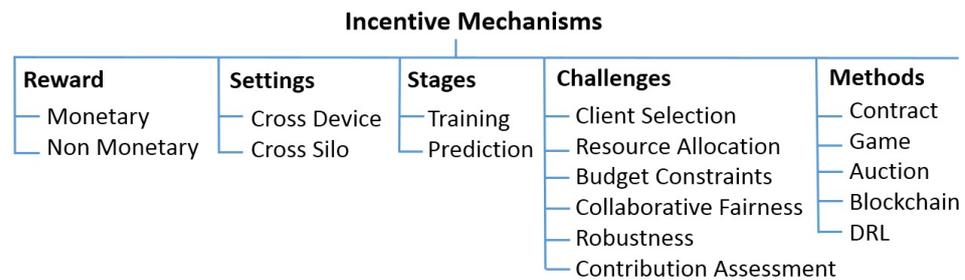


Figure 2. Taxonomy of incentive mechanisms in FL.

3.1. Reward

3.1.1. Monetary Incentives

Monetary incentives deal with the direct distribution of payoffs to FL clients and tend to motivate them to actively participate. This leads to better model training and, thus, high-quality contributions. Data owners and users of FL final models are assumed to be separate entities where data owners care more about their economic preferences.

3.1.2. Non-Monetary Incentives

Non-monetary incentives motivate clients by distributing different FL models or satisfying psychological needs such as reputation or virtual credit. Based on their work quality, such rewards are suitable in scenarios (1) unavailability of monetary budget, (2) clients value the final model more than the monetary incentive.

3.2. Settings

Incentives can be applied to both cross-device and cross-silo FL scenarios. Cross-device FL comprises many data owners, such as mobile or edge devices. In contrast, cross-silo FL has a relatively small number of data owners, such as medical hospitals or financial companies that collaboratively train a global model [31]. Recently, studies [32,33] have addressed cross-silo FL. In this methodology, our incentive scheme is designed for cross-device FL.

3.3. Stages

Figure 1 shows that the complete FL process includes model training and prediction. The training aims to obtain a high-quality global model, while prediction focuses on the good test performance of the designed system.

3.4. Challenges

- Client management to select qualified workers to join and remain in the training process.
- Resource allocation for clients based on the amount of work and data quality.
- Contribution evaluation to measure the contribution of each participant.
- Budget constraints due to the time-consuming commercialization and training of models or unavailability.
- Collaborative fairness corresponding to participant rewards should fairly reflect different levels of contributions.
- Robustness to targeted and untargeted attacks by malicious workers.

3.5. Methods

3.5.1. Contract Theory

Like economics, an FL contract is designed between a server (employer) and data owners (employees). The server cannot always be assumed to know the data owner's resources, data size, and quality. To overcome this issue of information asymmetry, the studies in [20,21,34] adopted contract theory to attract high-quality users. Contract-based selection can enhance participation and model quality; however, it only focuses on optimizing the user's utility/reward.

3.5.2. Game Theory

Game theory involves strategic interactions between rational participants. The goal is to determine the optimal strategy for a client's payoff in relation to other clients' strategies. The Stackelberg game is a game theoretic model with interaction between a leader (server) and follower (client) [35]. The central server gives a set of rules for participants to devise optimal strategies. This iterates until reaching the Nash equilibrium, where no player can improve its payoff by changing their strategy. However, this method is inefficient amid information asymmetry, undesirable behaviors, and non-i.i.d. data to incorporate multiple factors to design incentive schemes.

3.5.3. Blockchain

Blockchain is a decentralized digital ledger in peer-to-peer networks that records each user's encrypted transaction. This allows the features of robust and tamper-proof FL; thus, modern studies adopted it to provide privacy-preserving FL [36,37]. This security advantage can be a transparent record environment to store candidates' information. However, scalability due to time delays can occur in the blockchain FL system.

3.5.4. Auction Theory

Auction theory is an effective mathematical tool for task allocation and cost pricing. Auctions can be classified into forward and reverse auctions. In a forward auction, a seller displays an item, and bidders place the bid price. For our FL incentive scheme, we use the reverse auction where the server acting as a buyer requests a required service/task, and data owners acting as potential sellers perform the task. Le et al. [38] adopted auctions in wireless networks considering computational resources. In general, an auction scheme not only guarantees truthfulness or efficiency but also reduces the computation latency.

3.5.5. Deep Reinforcement Learning

Deep reinforcement learning (DRL) under FL consists of multiple agents performing actions under an environment to maximize rewards and privacy [39]. Incomplete information about clients' decisions and contribution concerns create a challenge in formulating

an ideal scheme. Therefore, Zhang et al. [40] adopted DRL with game theory to compute optimal trade pricing and tackle information asymmetry. The authors in [41] designed a hierarchical game to maintain a trade-off between maximizing payment and minimizing the learning rate through DRL. It focused on reducing the training time and determining a strategy based on experience.

4. Materials and Methods

4.1. Proposed Mechanism (FRIMFL)

We consider a collaborative horizontal FL environment with many potential candidates and model owner. Candidates serve as data owners and can be IoT or edge devices. The model owner broadcasts a training task anytime and recruit candidates to participate. Interested data owners examine their data quality and quantity, weigh the cost prices with reward possibility, and formulate strategies to submit bid prices to the server. The server combines their bids and reputation status for client selection and pays for their rewards. Figure 3 illustrates the overall workflow of our architecture.

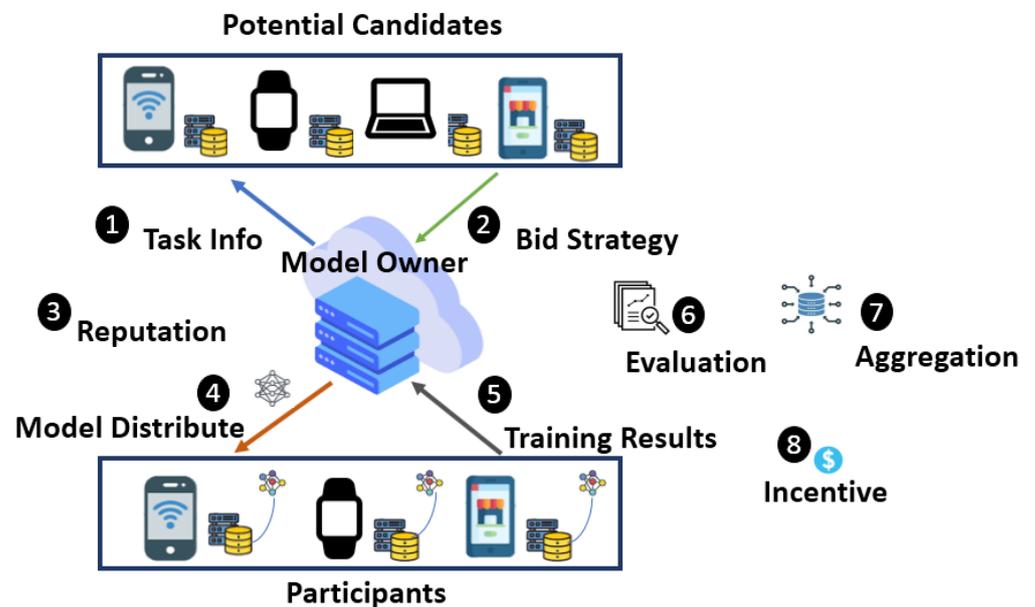


Figure 3. Architecture of FRIMFL.

The processes involved in each step are described below:

1. The server broadcasts the task information to N candidates, describing the budget and model requirements.
2. Interested candidates devise their bidding strategy based on data quantity or computational resources and submit their bid prices B_i^p to the server.
3. The server examines the candidates' reputation in all associated tasks and applies reverse auctions for global model distribution.
4. Selected participants from the candidates receive the initial global model and train local models iteratively on their local dataset.
5. The participants send their training results to the server.
6. The model owner collects training results, receives gradients, and executes quality detection through marginal loss evaluation.
7. The server aggregates quality models corresponding quality weights and measures participant contribution via a federated Shapley assessment and reputation to distribute payoffs.
8. Finally, participants are rewarded as per their level of reputation.

4.2. Reverse Auction-Based Optimal Client Selection

The assessment and selection of highly reputed candidates for any task are reflected by reverse auction design. The utility function of the auction is to maximize social surplus (SS) defined as a composite utility of all participants (P^u) and server (S^u) formulated as:

$$\max_{SS} \sum_{i=1}^N P_i^u + S^u, \quad (1a)$$

$$\text{s.t.} \quad \sum_{i=1}^N x_i p_i \leq B, \quad (1b)$$

$$x_i v_i \leq p_i. \quad (1c)$$

where x_i is a binary variable as if candidate i is selected to participate, $x_i = 1$, otherwise $x_i = 0$. The utility of participant i is expressed as the difference between its payoff incentive and cost calculated as $P_i^u = (p_i - v_i)$. After global aggregation, the server obtains a final global performance and distributes incentives. Thus, the server utility is calculated as $S^u = [B - \sum_{i=1}^N p_i]$. Here, Equation (1b) guarantees that the participant payoff p_i satisfies the budget feasibility, and Equation (1c) guarantees that the participant i should not be compensated less than its true cost v_i .

4.3. Design Properties

Since task requesters and clients are rational, FRIMFL must satisfy the following economic properties:

- Incentive Compatibility (IC) : The auction process satisfies incentive compatibility when all the participants obtain a maximum payoff by reporting bids truthfully.
- Individual Rationality (IR): When the participating users receive positive utility, the mechanism achieves incentive rationality.
- Budget Feasibility: The total incentive amount paid to participants does not exceed the model owners' budget.
- Computational Efficiency: The scheme can be computationally efficient if the winner determination and incentive distribution are computed within polynomial time.
- Aggregation Fairness: Each participant's aggregated weight shall correspond to its performance quality.
- Reward Fairness: Each participant shall be fairly rewarded, corresponding to their contribution levels for the task.

4.4. Quality Trust Assessment

With the ease of accessibility and lack of standard evaluation in the FL system, the participation influence cannot be fully ensured. The client's behavior can be classified as

- Positive Clients: These clients participate honestly, provide reliable model updates without malicious activity, and bid truthful data to complete training tasks.
- Negative Clients: These clients intend to have deceptive behavior, by either data poisoning through incorrect/sign-flip labels to decrease data accuracy or model poisoning to manipulate training performance [42,43].

An inherent way to achieve this is to evaluate the loss of the client model or update that exceeds a predetermined threshold [5]. But determining the threshold might be challenging where the performance of local models might improve in later training rounds if the threshold decreases, respectively. Therefore, we proposed a marginal error loss-based method executed by the server. A local model i is included in the global model m , and the loss l^m of this model is computed on a validation set. Likewise, this local model i is excluded, to make global model m' and loss $l_i^{m'}$ be computed. When a high-quality local model (M_i^l) is aggregated, the loss of the global model is reduced.

We define the threshold quality value ψ to accept the local model ($\psi_i < \psi$) as

$$\psi_i = l_i^{m'} - l^m. \quad (2)$$

In Vanilla FL [1], clients with large amounts of data receive higher weights. In FairFL [16], they receive equal weights. However, this might be an unfair way for an honest client to have a small amount of data with good quality. The existence of free riders can amplify the data size, and client heterogeneity can significantly affect the weights in model aggregation. Hence, it is necessary to aggregate local model weights passing the quality detection. Therefore, the aggregated weight (α_i) for participant i can be computed as

$$\alpha_i = \frac{\psi_i}{\sum_i \psi_i}. \quad (3)$$

Accordingly, the new aggregated global model (G_m) in round t can be determined as

$$G_m = \sum_i \alpha_i M_i^t. \quad (4)$$

The proposed weighted quality detection in FRIMFL is stated in Algorithm 1. Line 1 initializes an empty track of participants (R^t) kept by the model owner, followed by the loss calculation on the global model m . Lines (3–5) describe the evaluation process to compute for a marginal loss impact on all clients using Equation (2). Lines (6–12) present quality detection; for successful detection, (n_i^p) increases and vice versa. Finally, lines (13–16) describe the quality of the local weight used in the final global model aggregation.

Algorithm 1 FRIMFL quality detection

Input: R^t, ψ
Output: n_i^p, n_i^f, G_m

- 1: $R^t = \emptyset$
- 2: calculate l^m on global model m
- 3: **for** each client $i \in N$ **do**
- 4: calculate $l^{m'}$ on global model m'
- 5: calculate ψ_i using Equation (2)
- 6: **if** $\psi_i > \psi$ **then**
- 7: $R^t \leftarrow R^t \cup \{i\}$
- 8: $n_i^p ++$
- 9: **else**
- 10: $n_i^f ++$
- 11: **end if**
- 12: **end for**
- 13: **for** each client $i \in R^t$ **do**
- 14: calculate α_i using Equation (3)
- 15: **end for**
- 16: calculate G_m using Equation (4)
- 17: **return** n_i^p, n_i^f, G_m

4.5. Contribution Assessment

Fairness is an important parameter to evaluate contributions to attract and retain high-performing users in FL. We propose a per round evaluation method leverage through the Shapley value (SV) method [32,44]. It provides a fair reward distribution to participants as per their marginal contributions to calculate individual utility and encourages high-quality ones to join as early as possible. For instance, three clients, i.e., client 1, client 2, and client 3, collaborate on a task, given a characteristic evaluation function $V(\cdot)$, such that $V(1) = 40$, $V(2) = 60$, $V(3) = 80$, $V(1, 2) = 70$, $V(1, 3) = 75$, $V(2, 3) = 85$, and $V(1, 2, 3) = 90$. Table 1

depicts all possible combinations to quantify the contribution of all individuals. For any subset S , it is defined as

$$c_i = \frac{1}{N!} \sum_{S \subseteq N \setminus i} |S|!(N - |S| - 1)! [V(S \cup i) - V(S)]. \tag{5}$$

Thus, the collective contribution by participant i from round 1 to T can be expressed as

$$c_i = \sum_{t=1}^T c_i^t. \tag{6}$$

The contribution quantification in FL is coherent with the following properties.

- **Fairness:** Participants with similar models or updates shall receive similar contribution values. The contribution scale is correlated with the reward.
- **Availability:** The contribution value by negative clients shall be 0, as they have no impact on the global model in the current round.
- **Additivity:** With each cycle of global updates, both long- and short-term contributions are additive to the overall FL process.

We analyze fairness using the Pearson correlation coefficient. Let contributions on x axis such as $x = \{c_1, c_2, \dots, c_n\}$, rewards on y axis such as $y = \{p_1, p_2, \dots, p_n\}$ and average of x and y as $\{\hat{x}, \hat{y}\}$. The fairness correlation is given as

$$f_c = \frac{\sum_{i=1}^n (x_i - \hat{x})(y_i - \hat{y})}{n s_x s_y}, \tag{7}$$

where $\{s_x, s_y\}$ denote the standard deviations of x and y , respectively. The range of the correlation coefficient is $[-1, 1]$. The larger f_c value suggests high fairness (positive correlation), whereas the negative coefficient implies unfairness (negative correlation). The rewards for quality workers are positively correlated with contributions and reputation. If worker i performs higher than worker j ($c_i > c_j$), then $p_i > p_j$. The analysis and evaluation of f_c are discussed in the following sections.

Table 1. An example of the contribution evaluation in different coalitions.

Client	1-2-3	1-3-2	2-1-3	2-3-1	3-1-2	3-2-1	SV
1	40	40	10	5	0	5	16.67
2	30	15	60	60	10	5	30
3	20	35	20	25	80	80	20

4.6. Reputation Measurement

A reputation mechanism indirectly reflects a participant’s credibility and reliability leveraged through client selection and the reward distribution mechanisms. The effective and accurate reputation calculation is crucial for trustworthy FL because high-reputation clients with high performances play a crucial role in model training. The model owner manages the reputation of each participant i based on its contribution and the trust notion associated with quality detection. The trust value for each participant i is computed as

$$t_i = \frac{\omega n_i^p}{\omega n_i^p + (1 - \omega) n_i^f}. \tag{8}$$

where ω denotes the weight on the interaction events. The positive events n_i^p increase the trust score of the participants, and vice versa. To refrain from quality evaluation failure, negative events n_i^f are given a higher weight than positive events. Combining Equations (6) and (8), the reputation of any participant in round t can be stated as

$$r_i = t_i c_i. \quad (9)$$

4.7. Client Selection Reward Module

To achieve any desirable task, the model owner can select different types of participants based on their data quality, bid price, and performance. The participant's reward density is computed as

$$r_i^d = B_i^p / r_i. \quad (10)$$

Hence, to pick the early contributors (highly reputed), the participant set is arranged in the order of increasing reward density as

$$r_1^d < r_2^d < \dots < r_n^d. \quad (11)$$

The incentive price p_i for different quality participants in the ranking is

$$p_i = \begin{cases} r_i^d + 1, & \text{if } r_i \geq 0.85 \quad (\text{high reputed}); \\ r_i^d + 0.5, & \text{if } 0.70 < r_i < 0.85 \quad (\text{moderate high reputed}); \\ r_i^d, & \text{if } 0.6 < r_i < 0.70 \quad (\text{moderate low reputed}); \\ 0, & \text{if } r_i \leq 0.6 \quad (\text{low reputed}). \end{cases} \quad (12)$$

This shows that reputation performance can add an extra incentive of 1 or 0.5 to the user's overall payoff (Algorithm 2).

Algorithm 2 FRIMFL incentive allocation.

Input: $B, x_i, w, P = \{P_1, P_2, \dots, P_n\}$

Output: p_i

```

1:  $p_i = 0, x_i = 1$ 
2: for each client  $i \in P$  do
3:   Calculate  $c_i$  using Equation (6)
4:   Calculate  $t_i$  using Equation (8)
5:   Calculate  $r_i$  using Equation (9)
6:   Calculate  $r_i^d$  using Equation (10)
7:   Sort all using Equation (11)
8:   while  $\sum_{i=1}^n x_i p_i \leq B$  do
9:     Calculate  $p_i$  using Equation (12)
10:  end while
11: end for
12: return  $p_i$ 

```

5. Theoretical Analysis

Theorem 1 (FRIMFL achieves aggregation fairness). *FRIMFL gives the aggregation weight to each participant update based on its performance corresponding quality.*

Proof. The local model weight passing the quality detection is computed based on the performance instead of the amount of data, i.e., $\alpha_i = \frac{\psi_i}{\sum_i \psi_i}$. \square

Theorem 2 (FRIMFL achieves reward fairness). *FRIMFL fairly distributes rewards to individuals based on their performance contribution.*

Proof. To quantify fairness, f_c is $\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{ns_x s_y}$. For performers, bringing $y_i = p_i = t_i \frac{c_i}{\sum c_j} = t_i \frac{x_i}{\sum x_j}$ and assuming similar trust values $t_i = t_j$ into Equation (8), we obtain

$$\begin{aligned} f_c &= \frac{\sum_{i=1}^n (x_i - \frac{1}{n} \sum x_i) (t_i \frac{x_i}{\sum x_j} - \frac{1}{n} \sum t_i \frac{x_i}{\sum x_j})}{ns_x s_y} \\ &= \frac{\sum_{i=1}^n x_i (t_i \frac{x_i}{\sum x_j}) - x_i (\frac{1}{n} \sum t_i \frac{x_i}{\sum x_j}) - \frac{1}{n} \sum x_i (t_i \frac{x_i}{\sum x_j}) + \frac{1}{n} \sum x_i (\frac{1}{n} \sum t_i \frac{x_i}{\sum x_j})}{ns_x s_y} \\ &= \frac{\frac{t_i}{\sum x_j} \sum_{i=1}^n (x_i^2 - \frac{1}{n} \sum x_i x_i - \frac{1}{n} \sum x_i x_i + \frac{1}{n} \sum x_i \frac{1}{n} \sum x_i)}{ns_x (t_i \frac{x}{\sum x_j})} \\ &= \frac{\frac{t_i}{\sum x_j} \frac{1}{n} \sum_{i=1}^n (x_i - \frac{1}{n} \sum x_i)^2}{ns_x (t_i \frac{x}{\sum x_j})} \\ &= 1. \end{aligned}$$

□

Theorem 3 (FRIMFL satisfies individual rationality). *Each individual can receive positive utility.*

Proof. The utility for a selected individual i is computed as $u_i = x_i(p_i - v_i)$. If unselected, its utility is zero. Since the incentive is proportional to the contribution, $u_i \geq 0$. □

Theorem 4 (FRIMFL is budget feasible). *The total amount paid to the participating users does not exceed the server's budget.*

Proof. The scheme determines and rewards n participants subjected to $\sum_{i=1}^n x_i p_i \leq B$. For the winner i , the incentive is p_i , and for the loser i , the incentive is 0. □

6. Results

6.1. Experimental Settings

We evaluated the performance of the proposed FRIMFL incentive scheme using MNIST [45], Fashion MNIST [46], and CIFAR10 [47] image datasets. The MNIST dataset consists of handwritten digits [0–9], with 60,000 training and 10,000 test samples. Each image has 28×28 pixels, and the pixel channel is grayscale. The Fashion MNIST dataset consists of clothing images with 70,000 samples belonging to 10 categories. The image pixel size, channel, and division of training and test set are identical to the MNIST. The CIFAR10 dataset comprises different object images with 60,000 samples and 32×32 pixels belonging to 10 categories. For i.i.d. settings, the dataset is uniformly distributed among participants, each with equal random samples, denoted as (UNI). The dataset is randomly partitioned among participants for non-i.i.d. settings and client heterogeneity. To illustrate, five participants own {50, 100, 150, 200, 250} samples, denoted as (IMB). The model comprises a simple convolutional neural network (CNN) with two fully connected layers. Table 2 shows the parameter settings. We introduced negative or malicious behaviors in the FL system as

- Poison clients [42]: They perform training with some percentage of incorrect noisy labels to represent a degree of unreliability.
- Sign-flip clients [48]: They perform training with some percentage of flip labels to represent a degree of unreliability.

We compare our proposed FRIMFL mechanism with the following baselines.

- VFL [1]: It performs standard vanilla FL to randomly select fraction n individuals and calculates aggregation weights based on local dataset sizes.

- FairFL [16]: It gives equal weights to all clients for aggregating local models.
- Greedy: A mechanism always prefers candidates with lower bid prices only. It does not contain reputation and contribution assessment methods.
- RRAFL [17]: An auction and reputation (RRAFL)-based incentive scheme, primarily suited to uniform settings only. The reputation, quality-detection, and reward distributions are different from FRIMFL.

We compare the performances regarding the reverse auction, reputation-based selection, model convergence accuracy, robustness, and contribution fairness. The performance of the reverse auction shows the economic benefits and checks whether the proposed scheme satisfies IR and IC design properties. The performance of reputation-based selection shows how well the scheme can select participants with good data quality. The performance of model accuracy shows the convergence speed and the stability of the scheme. The robustness through cumulative rewards reflects the effectiveness of quality-reputation-based selection to promote positive participation and eliminate negative participation. Lastly, the performance of contribution fairness shows how well the scheme promotes contributions from individual participants.

Table 2. Parameter settings for simulation results.

Parameter	Value
Number of participants (n)	5–20
Bid price (B_i^p)	5–10
Budget (B)	100–300
Learning rate (η)	0.05
Batch size (B)	100
Loss quality threshold (ψ)	−0.03

6.2. Performance of Reverse Auction

Figure 4 compares the three schemes in terms of economic perspective. With 20 participants and a budget of 250, FRIMFL distributes higher payoffs and achieves a 7% higher surplus than RRAFL and Greedy. This is due to an increase in the selection probability of clients and their willingness to participate. Similarly, we compare the payments and cost value sum of all clients to join the training, and the results show that FRIMFL satisfies IR and IC design properties. This provides significant economic benefit for users to participate and makes the overall trade market more stable than both Greedy and RRAFL.

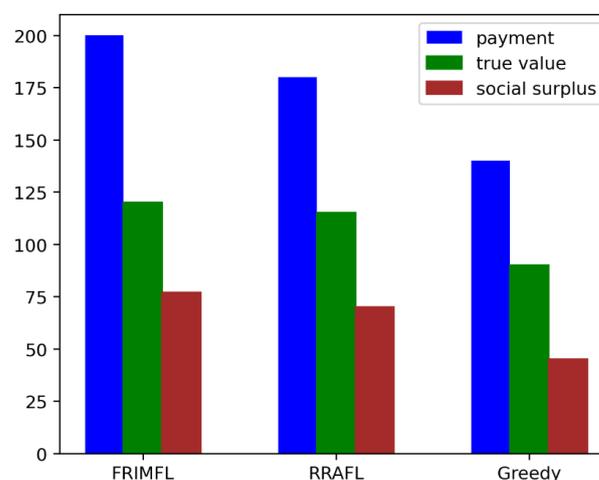


Figure 4. Comparison of economic benefit.

6.3. Performance of Reputation-Based Selection

We evaluate the reputation performance of truthful participants with different data quality rates. Table 3 shows the impact of the data accuracy level (p) on the average reputation for both schemes. Assuming honest participants with at least 60% true quality data, we can observe that the reputation reduces as the individual data accuracy rate decreases with non-linearity. For example, for the clients with a quality rate of 1 (i.e., a dataset containing all accurate labels) and a quality rate of 0.8 (i.e., a dataset containing 80% accurate labels and 20% incorrect labels), FRIMFL gives a higher reputation than RRAFL to those participants with higher quality rates. The reputation of FRIMFL is relatively higher than that of RRAFL due to the fair estimation of contributions. This shows that clients with good-quality data can receive a higher reputation for better incentives.

Table 3. Reputation analysis with different data quality rates.

	1		0.8		0.7		0.6	
Dataset	RRAFL	FRIMFL	RRAFL	FRIMFL	RRAFL	FRIMFL	RRAFL	FRIMFL
MNIST	0.965	0.974	0.698	0.742	0.588	0.635	0.431	0.508
CIFAR10	0.951	0.959	0.638	0.711	0.504	0.601	0.416	0.494
FMNIST	0.962	0.970	0.664	0.737	0.581	0.640	0.425	0.510

The evaluation of the model accuracy under the existence of different reputed clients is illustrated in Figure 5. For datasets [MNIST, CIFAR10, FMNIST], the highest accuracies [0.979, 0.801, 0.981] are observed in the presence of highly reputed (i.e., positive) clients, whereas the lowest accuracies [0.914, 0.722, 0.920] were reported with low reputed clients after 10 global epochs. The low-reputed participants had greater influence than the medium reputed participants on model convergence as they exhibit negative behavior by the falsified quality and lesser contribution towards the global model update. This also shows that slow model convergence is not guaranteed after 10 epochs and could create a loss to the model owner regarding payoff and system performance. With positive and medium clients, model convergence is not only fast but overall useful for model owners.

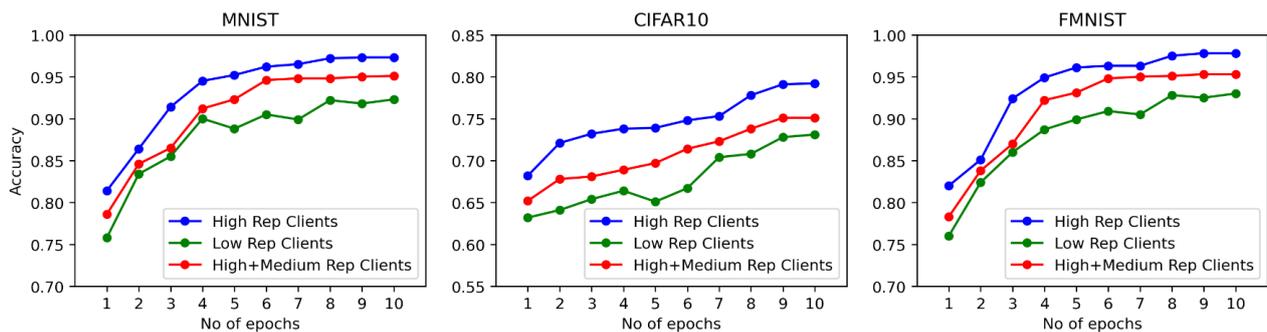


Figure 5. Effect on model accuracy by different reputed clients.

Table 4 shows the percentage of reliable participants selected in all four schemes on three datasets. Overall, FRIMFL selects the most participants, which means that it has effectively encouraged participation. For example, in CIFAR10, the proportion of client selection with good data quality is 0.94 in FRIMFL, significantly higher than 0.92 in RRAFL, 0.62 in VFL, and 0.40 in Greedy. Therefore, FRIMFL can benefit the model owners by providing higher chances to select users with better data quality.

Table 4. Participant selection probability in all four schemes.

Dataset	VFL	Greedy	RRAFL	FRIMFL
MNIST	0.61	0.41	0.93	0.95
CIFAR10	0.62	0.40	0.92	0.94
FMNIST	0.61	0.41	0.92	0.95

Figure 6 compares the accuracy of the final aggregated model in all four schemes. FRIMFL not only filters malicious behaviors through quality detection but also fairly aggregates weights to enhance the model accuracy than all three FairFL, RRAFL, and VFL at the end of task completion. Among the baselines, FairFL achieves a relatively higher performance in terms of convergence and stability than both RRAFL and VFL. This suggests that aggregation fairness handling data heterogeneity can make the global model more accurate and indicates that all reliable participants selected by FRIMFL have a positive effect on the global model accuracy, reaching [0.981, 0.812, 0.988] in three datasets [MNIST, CIFAR10, FMNIST], respectively, at the earliest.

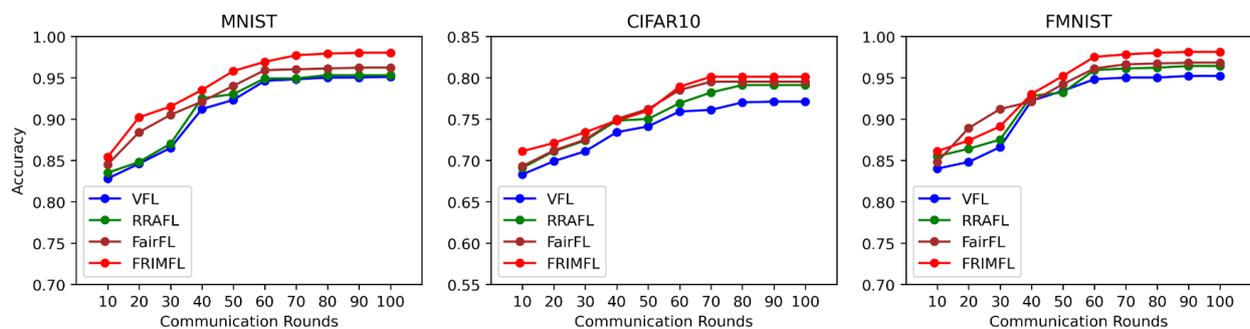


Figure 6. Performance of model accuracy.

The advantage of FRIMFL in an unreliable federation is shown in Figure 7. The x axis represents communication iterations, and the y axis represents the global model accuracy. The percentage of sign-flip and poison clients is equal to 75% to demonstrate untrustworthiness. The combination of such clients causes damage to the global model on all three datasets. The malicious behavior not only slows down the convergence speed but also requires three times more iterations to reach a stable accuracy of 0.812. In contrast, the honest behavior achieves a convergence accuracy of 0.963 approximately at 20 communication iterations. Therefore, the damage caused by malicious activity increases with a higher proportion of faulty labels and training updates.

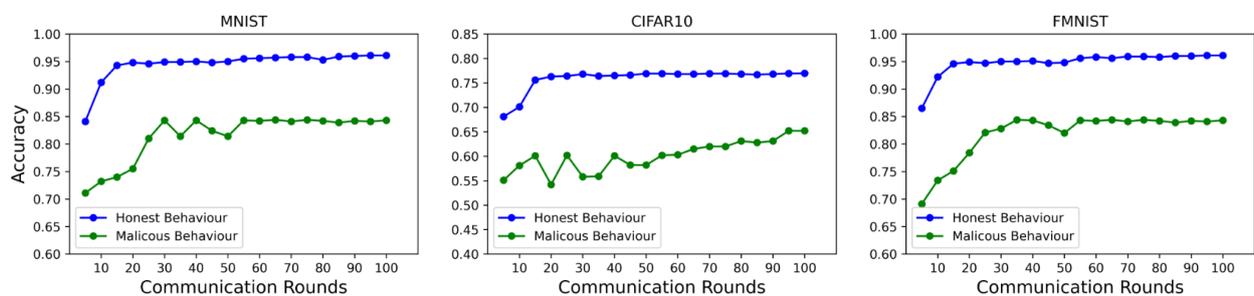


Figure 7. Model convergence under client behaviours.

6.4. Performance of Model Accuracy

Figure 8 shows the individual accuracy performance by five data owners in each training round for all three datasets. The participant’s local accuracy rate is normal as they join the training coalition with a true data quality. With positive behavior and reliable

updates, the server detects and compares local models for global model aggregation. It can be seen that the local accuracy saturates after five communication rounds.

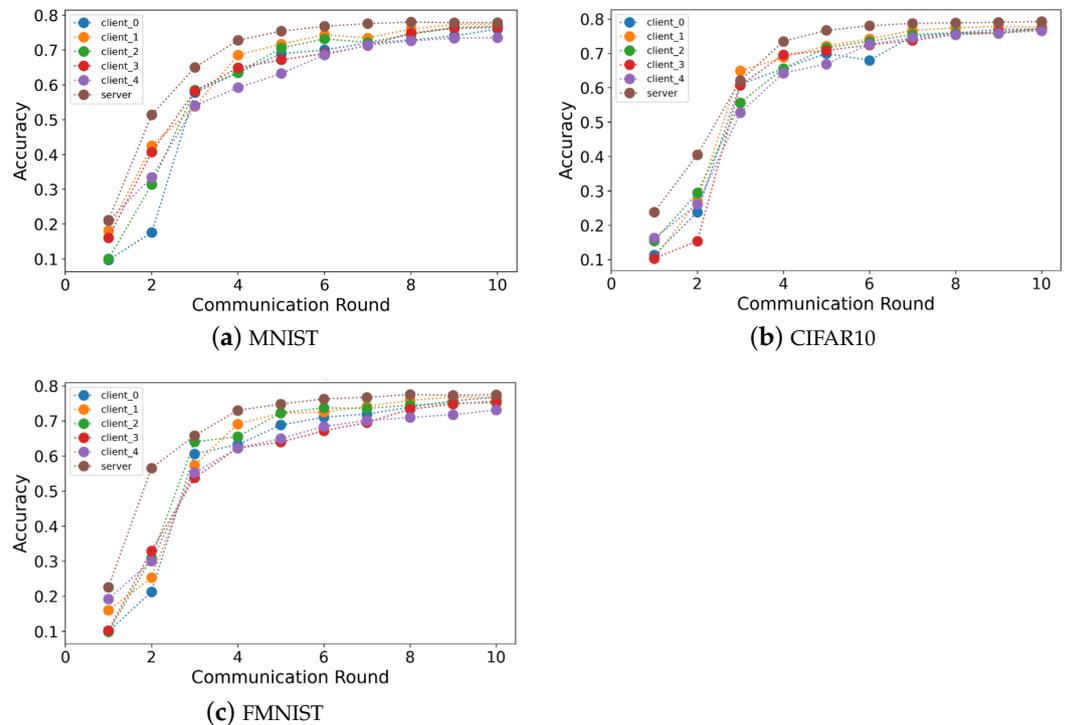


Figure 8. Performance of accuracy by 5 different clients.

6.5. Performance of Contribution Fairness

Figure 9 shows the individual contribution level by five data owners in each training round for all three datasets. Initially, the contributions were different for all clients. At training round 3, the highest contributions were observed. Client 0 appears to be the highest in all three datasets, whereas clients 3 and 4 contributed the least. Similar influence by participants in later training rounds demonstrates that FRIMFL promotes early contributions by all data owners.

For three datasets, we compare FRIMFL with the RRAFL baseline in terms of fairness measured through the Pearson correlation coefficient, as shown in Table 5. The highest f_c was obtained in uniform settings (UNI) in both mechanisms. The performance of RRAFL is lower because the contribution measurement via cosine similarity is not ideal for capturing data heterogeneity where clients might provide similar updates to improve their performance.

Table 5. Calculation of Pearson correlation coefficient.

	MNIST		CIFAR10		FMNIST	
Scheme	UNI	IMB	UNI	IMB	UNI	IMB
RRAFL	0.989	0.972	0.979	0.960	0.987	0.973
FRIMFL	0.992	0.988	0.985	0.973	0.993	0.986

The robustness of FRIMFL in fair reward distribution is shown in Figure 10. The x axis represents communication iterations and the y axis represents participants' cumulative incentive, where pf is the proportion of data owners' unreliable (poison-flip) data. The reward values vary with different pf values. Reputed clients with quality data (low pf) are considered collaborators and thus receive rewards. This shows that rewards are positively related to quality and contributions. Clients with malicious data (high pf) who have

malicious interactions (bad reputation) are considered to be negative participants and hence punished by the server for obtaining rewards.

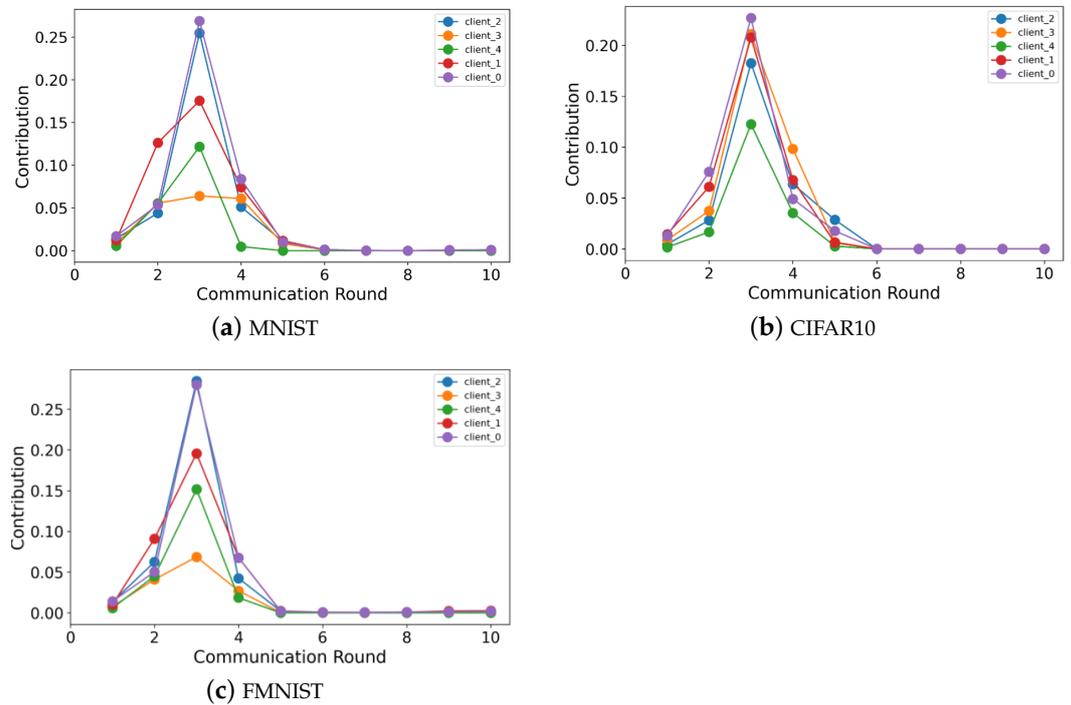


Figure 9. Performance of contribution by 5 different clients.

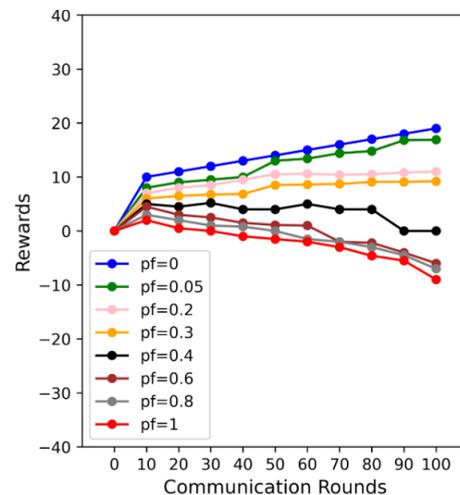


Figure 10. Cumulative rewards of participants with different data.

7. Discussion

The social surplus is relatively higher in the proposed mechanism due to increased participants' desire to select reverse auction. The cost–benefit makes the FL service more active. The Shapley method allows the measuring of client contributions and their impact on the overall performance of the model. Our mechanism implicitly assumes a predetermined incentive budget. However, this might not hold in practice in later training periods [12]. As for future works, a reimbursement option with dynamic incentivizing for long-term participation in diverse scenarios could be investigated. Additionally, the budget restriction only applies to the payments that the system employs to hold truthfulness. The convergence of the model is affected by the influence of negative/malicious clients since they produce

misleading gradients. Spoofing attacks with robust defense can also be explored to complete the definition of security in incentive schemes.

8. Conclusions

In this paper, we proposed a fair and reliable incentive mechanism (FRIMFL) for federated learning services between FL model owners and heterogeneous clients. We designed our scheme considering client selection and the reward fairness of truthful clients. First, we adopted reverse auctions to estimate the model training costs and motivate clients to participate. Second, we introduced a reputation mechanism to detect and differentiate client behaviors. Participants' reputation is formulated by combining trust via quality detection and the contribution evaluation via the Shapley method. Finally, by integrating the reverse auction with reputation, we designed a selection and reward mechanism to distribute incentives to individuals for their overall performances. FRIMFL fairly allocates the incentive reward to positive participants and prevents negative ones from damaging the quality model. Through the theoretical analysis, we proved that FRIMFL satisfies individual rationality, incentive compatibility, aggregation fairness, and budget feasibility. Experimental results have shown that our scheme selects the best combination of clients to maximize the social surplus in the FL trading marketplace. The proposed FRIMFL can significantly enhance the economic benefit by 7% based on higher client selection and the model accuracy of FL tasks from 0.9451 to 0.9832 with faster convergence, hence encouraging active participation for the stable development of the FL ecosystem and thus can be a feasible option to incentivize clients.

Author Contributions: Conceptualization, A.A. and B.J.C.; Methodology, A.A.; Software, A.A.; Validation, A.A.; Formal Analysis, A.A.; Investigation, A.A. and B.J.C.; Resources, A.A.; Data Curation, A.A.; Writing—Original Draft Preparation, A.A.; Writing—Review and Editing, B.J.C.; Visualization, A.A.; Supervision, B.J.C.; Project Administration, B.J.C.; Funding Acquisition, B.J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the MSIT Korea under the NRF Korea (NRF-2022R1A2C4001270), the Innovative Human Resource Development for Local Intellectualization support program (IITP-2023-RS-2022-00156360) supervised by the IITP, and the KIAT grant funded by the Korean government (MOTIE) (P0017123, The Competency Development Program for Industry Specialist). This research was also supported by the National Research Foundation (NRF) of Korea (NRF-2020K1A3A1A68093469), funded by the Ministry of Science and ICT (MSIT) Korea, and by the Department of Biotechnology (India) (DBT/IC-12031(22)-ICD-DBT).

Data Availability Statement: MNIST, FMNIST, and CIFAR10 datasets used in this article are publicly available and can be found at [45–47]. The code used to evaluate FRIMFL is available at <https://github.com/Abrar-Ahmed-96/incentivefl>, accessed on 24 July 2023.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

α_i	Local model weight by participant i
B	Federation budget
B_i^p	Bid price of participant i
c_i	Contribution of participant i
CNN	Convolutional neural network
DRL	Deep reinforcement learning
f_c	Fairness correlation coefficient
FL	Federated learning
IC	Incentive compatibility
IR	Individual rationality
l_i^m	Loss of model with participant i
$l_i^{m'}$	Loss of model without participant i

n_i^p	Number of passing detections for participant i
n_i^f	Number of failing detections for participant i
p	Data quality rate
p_i	Payoff of participant i
P_i^u	Utility of participant i
R^t	Record of quality detection
r_i^d	Reward density of participant i
r_i	Reputation of participant i
S_i^u	Server i utility
SS	Social surplus
SV	Shapley value
t_i	Trust on participant i
x_i	Selection flag for participant i

References

- McMahan, H.B.; Moore, E.; Ramage, D.; Arcas, B.A.Y. Federated Learning of Deep Networks using Model Averaging. *arXiv* **2016**, arXiv:1602.05629.
- Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated Machine Learning: Concept and Applications. *ACM Trans. Intell. Syst. Technol.* **2019**, *10*, 1–19. [[CrossRef](#)]
- Zhan, Y.; Zhang, J.; Hong, Z.; Wu, L.; Li, P.; Guo, S. A Survey of Incentive Mechanism Design for Federated Learning. *IEEE Trans. Emerg. Top. Comput.* **2022**, *10*, 1035–1044. [[CrossRef](#)]
- Yang, T.; Andrew, G.; Eichner, H.; Sun, H.; Li, W.; Kong, N.; Ramage, D.; Beaufays, F. Applied Federated Learning: Improving Google Keyboard Query Suggestions. *arXiv* **2018**, arXiv:1812.02903.
- Shayan, M.; Fung, C.; Yoon, C.J.M.; Beschastnikh, I. Biscotti: A Ledger for Private and Secure Peer-to-Peer Machine Learning. *arXiv* **2018**, arXiv:1811.09904.
- Yu, H.; Liu, Z.; Liu, Y.; Chen, T.; Cong, M.; Weng, X.; Niyato, D.T.; Yang, Q. A Fairness-aware Incentive Scheme for Federated Learning. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, New York, NY, USA, 7–8 February 2020.
- Zhou, Z.; Liu, P.; Feng, J.; Zhang, Y.; Mumtaz, S.; Rodriguez, J. Computation Resource Allocation and Task Assignment Optimization in Vehicular Fog Computing: A Contract-Matching Approach. *IEEE Trans. Veh. Technol.* **2019**, *68*, 3113–3125. [[CrossRef](#)]
- Dinh, C.T.; Tran, N.H.; Nguyen, M.N.H.; Hong, C.S.; Bao, W.; Zomaya, A.Y.; Gramoli, V. Federated Learning Over Wireless Networks: Convergence Analysis and Resource Allocation. *IEEE/ACM Trans. Netw.* **2021**, *29*, 398–409. [[CrossRef](#)]
- Yang, D.; Xue, G.; Fang, X.; Tang, J. Incentive Mechanisms for Crowdsensing: Crowdsourcing with Smartphones. *IEEE/ACM Trans. Netw.* **2016**, *24*, 1732–1744. [[CrossRef](#)]
- Biggio, B.; Nelson, B.; Laskov, P. Poisoning Attacks against Support Vector Machines. In Proceedings of the ICML, Edinburgh, UK, 26 June–1 July 2012.
- Lin, J.; Du, M.; Liu, J. Free-riders in Federated Learning: Attacks and Defenses. *arXiv* **2019**, arXiv:1911.12560.
- Kairouz, P.; McMahan, H.B.; Avent, B.; Bellet, A.; Bennis, M.; Nitin Bhagoji, A.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. Advances and Open Problems in Federated Learning. *Found. Trends Mach. Learn.* **2021**, *14*, 1–210. [[CrossRef](#)]
- Zhang, J.; Shu, Y.; Yu, H. Fairness in Design: A Framework for Facilitating Ethical Artificial Intelligence Designs. *Int. J. Crowd Sci.* **2023**, *7*, 32–39. [[CrossRef](#)]
- Nishio, T.; Yonetani, R. Client Selection for Federated Learning with Heterogeneous Resources in Mobile Edge. In Proceedings of the ICC 2019—2019 IEEE International Conference on Communications (ICC), Shanghai, China, 20–24 May 2019; pp. 1–7. [[CrossRef](#)]
- Li, T.; Hu, S.; Beirami, A.; Smith, V. Ditto: Fair and Robust Federated Learning Through Personalization. In Proceedings of the International Conference on Machine Learning, Virtual Event, 13–18 July 2020.
- Michieli, U.; Ozay, M. Are All Users Treated Fairly in Federated Learning Systems? In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021; pp. 2318–2322. [[CrossRef](#)]
- Zhang, J.; Wu, Y.; Pan, R. Incentive Mechanism for Horizontal Federated Learning Based on Reputation and Reverse Auction. In Proceedings of the Web Conference 2021 (WWW '21), Ljubljana, Slovenia, 19–23 April 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 947–956. [[CrossRef](#)]
- Shi, Y.; Yu, H.; Leung, C. A Survey of Fairness-Aware Federated Learning. *arXiv* **2021**, arXiv:2111.01872.
- Zeng, R.; Zeng, C.; Wang, X.; Li, B.; Chu, X. A Comprehensive Survey of Incentive Mechanism for Federated Learning. *arXiv* **2021**, arXiv:2106.15406.
- Kang, J.; Xiong, Z.; Niyato, D.T.; Xie, S.; Zhang, J. Incentive Mechanism for Reliable Federated Learning: A Joint Optimization Approach to Combining Reputation and Contract Theory. *IEEE Internet Things J.* **2019**, *6*, 10700–10714.

21. Kang, J.; Xiong, Z.; Niyato, D.; Yu, H.; Liang, Y.C.; Kim, D.I. Incentive Design for Efficient Federated Learning in Mobile Networks: A Contract Theory Approach. In Proceedings of the 2019 IEEE VTS Asia Pacific Wireless Communications Symposium (APWCS), Singapore, 28–30 August 2019; pp. 1–5. [\[CrossRef\]](#)
22. Ye, D.; Yu, R.; Pan, M.; Han, Z. Federated Learning in Vehicular Edge Computing: A Selective Model Aggregation Approach. *IEEE Access* **2020**, *8*, 23920–23935. [\[CrossRef\]](#)
23. Ezzeldin, Y.H.; Yan, S.; He, C.; Ferrara, E.; Avestimehr, A.S. FairFed: Enabling Group Fairness in Federated Learning. *Proc. AAAI Conf. Artif. Intell.* **2023**, *37*, 7494–7502. [\[CrossRef\]](#)
24. Horváth, S.; Laskaridis, S.; Almeida, M.; Leontiadis, I.; Venieris, S.; Lane, N. FjORD: Fair and Accurate Federated Learning under heterogeneous targets with Ordered Dropout. In *Advances in Neural Information Processing Systems*; Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 12876–12889.
25. Zhang, J.; Li, C.; Robles-Kelly, A.; Kankanhalli, M. Hierarchically Fair Federated Learning. *arXiv* **2020**, arXiv:2004.10386.
26. Lyu, L.; Yu, J.; Nandakumar, K.; Li, Y.; Ma, X.; Jin, J.; Yu, H.; Ng, K.S. Towards Fair and Privacy-Preserving Federated Deep Models. *IEEE Trans. Parallel Distrib. Syst.* **2019**, *31*, 2524–2541. [\[CrossRef\]](#)
27. Thi Le, T.H.; Tran, N.H.; Tun, Y.K.; Nguyen, M.N.H.; Pandey, S.R.; Han, Z.; Hong, C.S. An Incentive Mechanism for Federated Learning in Wireless Cellular Networks: An Auction Approach. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 4874–4887. [\[CrossRef\]](#)
28. Cong, M.; Yu, H.; Weng, X.; Qu, J.; Liu, Y.; Yiu, S.M. A VCG-based Fair Incentive Mechanism for Federated Learning. *arXiv* **2020**, arXiv:2008.06680.
29. Zeng, R.; Zhang, S.; Wang, J.; Chu, X. FMore: An Incentive Scheme of Multi-dimensional Auction for Federated Learning in MEC. In Proceedings of the 2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS), Singapore, 29 November–1 December 2020; pp. 278–288. [\[CrossRef\]](#)
30. Rehman, M.H.; Salah, K.; Damiani, E.; Svetinovic, D. Towards Blockchain-Based Reputation-Aware Federated Learning. In Proceedings of the IEEE INFOCOM 2020—IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Toronto, ON, Canada, 6–9 July 2020; pp. 183–188.
31. Li, T.; Sahu, A.K.; Talwalkar, A.S.; Smith, V. Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Process. Mag.* **2020**, *37*, 50–60. [\[CrossRef\]](#)
32. Pandey, S.R.; Nguyen, L.D.; Popovski, P. FedToken: Tokenized Incentives for Data Contribution in Federated Learning. *arXiv* **2022**, arXiv:2209.09775.
33. Deng, Y.; Lyu, F.; Ren, J.; Chen, Y.C.; Yang, P.; Zhou, Y.; Zhang, Y. FAIR: Quality-Aware Federated Learning with Precise User Incentive and Model Aggregation. In Proceedings of the IEEE INFOCOM 2021—IEEE Conference on Computer Communications, Vancouver, BC, Canada, 10–12 May 2021; pp. 1–10. [\[CrossRef\]](#)
34. Ding, N.; Fang, Z.; Huang, J. Incentive Mechanism Design for Federated Learning with Multi-Dimensional Private Information. In Proceedings of the 2020 18th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT), Volos, Greece, 15–19 June 2020; pp. 1–8.
35. Feng, S.; Niyato, D.T.; Wang, P.; Kim, D.I.; Liang, Y.C. Joint Service Pricing and Cooperative Relay Communication for Federated Learning. In Proceedings of the 2019 International Conference on Internet of Things (Things) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Halifax, NS, Canada, 30 July–3 August 2018; pp. 815–820.
36. Weng, J.; Weng, J.; Zhang, J.; Li, M.; Zhang, Y.; Luo, W. DeepChain: Auditable and Privacy-Preserving Deep Learning with Blockchain-Based Incentive. *IEEE Trans. Dependable Secur. Comput.* **2021**, *18*, 2438–2455. [\[CrossRef\]](#)
37. Bao, X.; Su, C.; Xiong, Y.; Huang, W.; Hu, Y. FLChain: A Blockchain for Auditable Federated Learning with Trust and Incentive. In Proceedings of the 2019 5th International Conference on Big Data Computing and Communications (BIGCOM), Qingdao, China, 9–11 August 2019; pp. 151–159. [\[CrossRef\]](#)
38. Le, T.H.T.; Tran, N.H.; Tun, Y.K.; Han, Z.; Hong, C.S. Auction based Incentive Design for Efficient Federated Learning in Cellular Wireless Networks. In Proceedings of the 2020 IEEE Wireless Communications and Networking Conference (WCNC), Seoul, Republic of Korea, 25–28 May 2020; pp. 1–6. [\[CrossRef\]](#)
39. Qi, J.; Zhou, Q.; Lei, L.; Zheng, K. Federated Reinforcement Learning: Techniques, Applications, and Open Challenges. *arXiv* **2021**, arXiv:2108.11887.
40. Zhan, Y.; Li, P.; Qu, Z.; Zeng, D.; Guo, S. A Learning-Based Incentive Mechanism for Federated Learning. *IEEE Internet Things J.* **2020**, *7*, 6360–6368. [\[CrossRef\]](#)
41. Zhan, Y.; Zhang, J. An Incentive Mechanism Design for Efficient Edge Learning by Deep Reinforcement Learning Approach. In Proceedings of the IEEE INFOCOM 2020—IEEE Conference on Computer Communications, Toronto, ON, Canada, 6–9 July 2020; pp. 2489–2498. [\[CrossRef\]](#)
42. Tolpegin, V.; Truex, S.; Gursoy, M.E.; Liu, L. Data Poisoning Attacks Against Federated Learning Systems. In Proceedings of the ESORICS, Guildford, UK, 14–18 September 2020.
43. Shejwalkar, V.; Houmansadr, A. Manipulating the Byzantine: Optimizing Model Poisoning Attacks and Defenses for Federated Learning. In Proceedings of the NDSS, Virtually, 21–25 February 2021.
44. Wang, G.; Dang, C.X.; Zhou, Z. Measure Contribution of Participants in Federated Learning. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 2597–2604. [\[CrossRef\]](#)

45. Deng, L. The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]. *IEEE Signal Process. Mag.* **2012**, *29*, 141–142. [[CrossRef](#)]
46. Xiao, H.; Rasul, K.; Vollgraf, R. Fashion-MNIST: A Novel Image Dataset for Benchmarking Machine Learning Algorithms. *arXiv* **2017**, arXiv:1708.07747.
47. Krizhevsky, A. *Learning Multiple Layers of Features from Tiny Images*; University of Toronto: Toronto, ON, Canada, 2009.
48. Fung, C.; Yoon, C.J.M.; Beschastnikh, I. The Limitations of Federated Learning in Sybil Settings. In Proceedings of the International Symposium on Recent Advances in Intrusion Detection, San Sebastian, Spain, 14–18 October 2020.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.